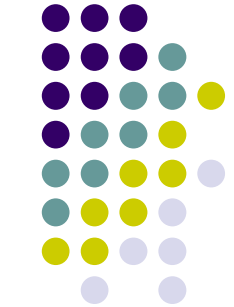
Collaborative Ontology Development in Protégé

Tania Tudorache
Stanford University

- Ontolog forum invited talk-

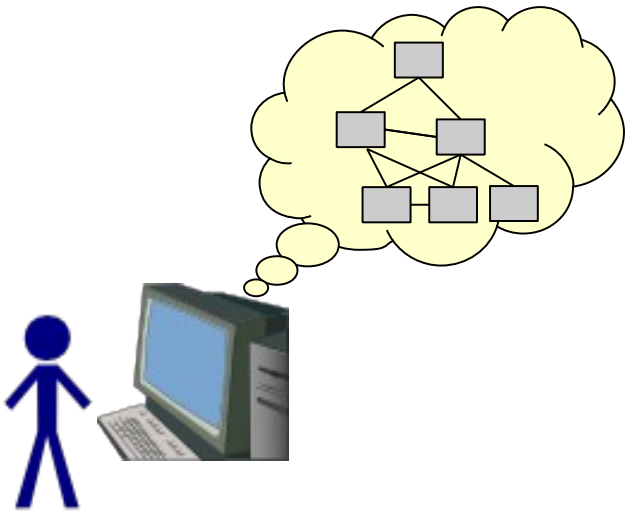
04. October 2007

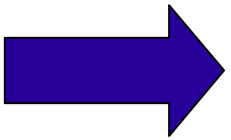
Outline

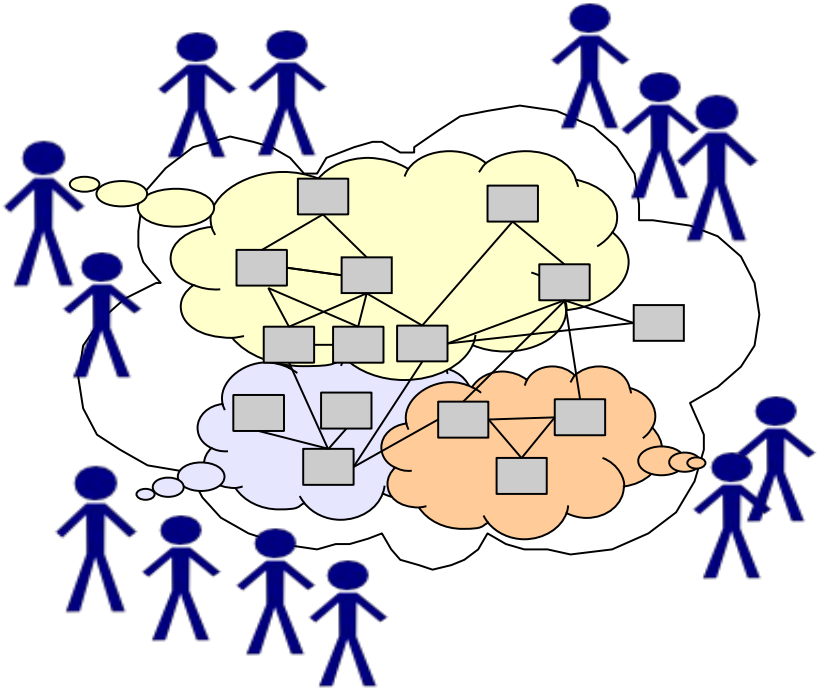


- Introduction and Background
- Tools for collaborative knowledge development
- Use cases
- Workflows
- Collaborative Protégé
- Future directions

Evolution of ontology development







- Single ontology developer
- Small size ontologies
- Desktop applications

- Community of ontology developers
- Large ontologies
- Web applications

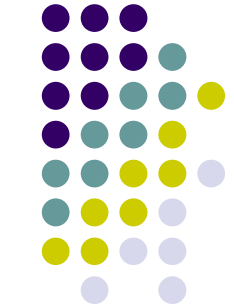
Collaborative ontology development

Two or more people interact and exchange knowledge to build a common ontology in pursuit of a shared, collective, bounded goal*

- Interaction may be indirect, but required
- Argumentation as a common interaction means
- Simple contribution not enough
- Bounded goal: beginning and end
- Collaborators may have individual goals

*Adapted from: http://collab.blueoxen.net/cgi-bin/wiki.pl?Collaboration

Outline

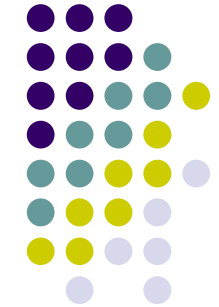


- Introduction and Background
- **Tools for collaborative knowledge development**
- Use cases
- Workflows
- Collaborative Protégé
- Future directions

Requirements

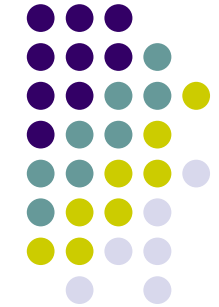
- Annotations of ontology components and changes
  - Marginal notes
  - Discussion threads
- Workflow support
  - Change proposals and voting
  - Rights management
  - Access privileges
- Views on the data
- Ways to establish trust, credibility
- Ratings and voting
- Comments and provenance of ratings and votes

Requirements (cont)



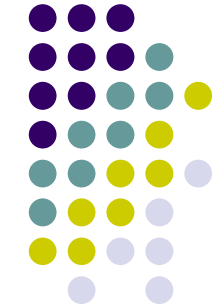
- A Web interface rather than an applet or an application
- Change history for a concept
- Ability to inform the user if something they posted was modified by someone else
- Undo and rollback based on change history
- Ability to query an old version through the vocabulary of the new one
- API access to changes; printed summary of changes
- Metrics attached to a concept

Tools for collaborative knowledge development



- Semantic wikis
  - *Semantic MediaWiki, BoWiki, etc.*
- Annotation of Web resources, tagging, bookmarks
  - *BibSonomy, SOLBOLEO*
- Ontology editors
  - *Collaborative Protégé, OntoWiki, Hozo*
- Domain-specific collaboration tools
  - *SWAN, Knewco, CBioC*

The CKC 2007 Challenge*



- At the Workshop on Social and Collaborative Construction of Structured Knowledge, Banff, Canada
- Goal: Find out what is the state of the art for the tools for collaborative construction of structured knowledge
- Get users to try out different tools
- Learn what users expect from such tools, what works and what doesn't
- It was not an evaluation of the tools themselves
- Tech Report available with challenge results

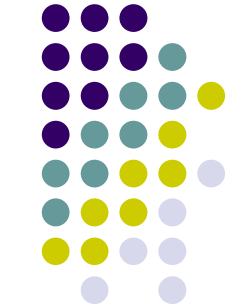
* http://km.aifb.uni-karlsruhe.de/ws/ckc2007/challenge.html

The CKC 2007 Challenge
Participating tools

- BibSonomy (University of Kassel, Germany)
- Collaborative Protégé (Stanford University, US)
- DBin (Universita Politecnica delle Marche, Italy)
- Hozo (University of Osaka, Japan)
- OntoWiki (University of Leipzig, Germany)
- SOBOLEO (Forschungszentrum Informatik, FZI, Germany)

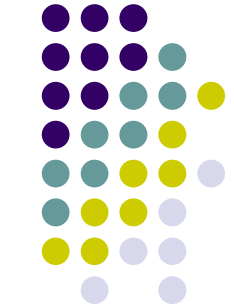
Different capabilities and focus

The CKC 2007 Challenge
Range of tools



- Annotation of Web resources, tagging, bookmarks
  - *BibSonomy*
  - *SOBOLEO*
- Ontology editors
  - *Collaborative Protégé*
  - *OntoWiki*
  - *Hozo*
- Tools with discussion and rating facilities
  - *Collaborative Protégé*
  - *DBin*

The CKC 2007 Challenge
Popular features



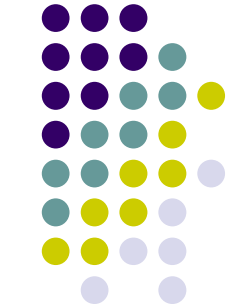
BibSonomy	postBookmark and postPublication buttons in a Browser; Upload to EndNote
Collaborative Protégé	Discussion, voting, chat Stable tool
DBin	Customizable UI
Hozo	Visualization
OntoWiki	Maps Ratings
SOBOLEO	Ease of use

See more results in the Tech Report

The CKC 2007 Challenge
Discussions

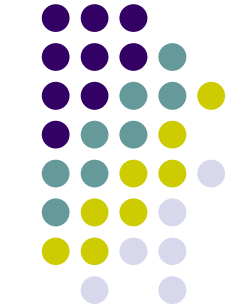
- Personal vs shared space
  - BibSonomy kept everyone's personal space separate
  - Other tools had everyone editing in the same shared space
  - Which model is more appropriate for ontology development?
- Can we even develop ontologies collaboratively?
  - *"I also think that collaborative ontology building could become very messy with a non-trivial user base; probably it would get on my nerves when people start shoving "my" concepts around."*
- What level of expressive power is appropriate?
  - Not supporting advanced OWL constraints -- is this a limitation?
- Collaborative workflows to achieve consensus

Outline



- Introduction and Background
- Tools for collaborative knowledge development
- **Use cases**
- Workflows
- Collaborative Protégé
- Future directions

Use cases



I. Ontology for Hospital Enterprise Architecture
   - Perot Systems

II. National Cancer Institute (NCI) Thesaurus
   - NCI Center for Bioinformatics

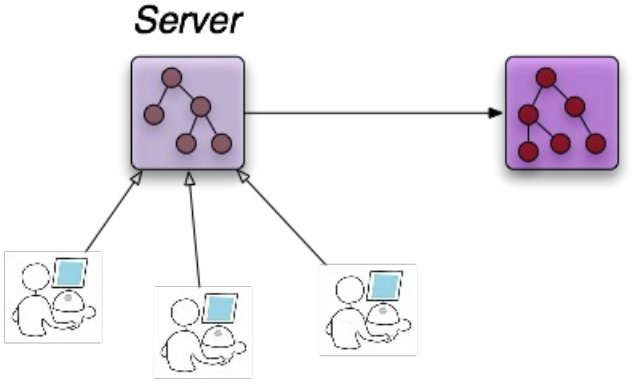
III. Open Biomedical Ontologies (OBO)

IV. International Classification of Diseases (ICD-10)

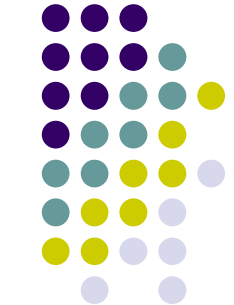
Case I: Multi-client editing

Perot Systems

- *Multiple users editing the same ontology at the same time*
- *All changes are seen immediately*
- *No conflict resolution*

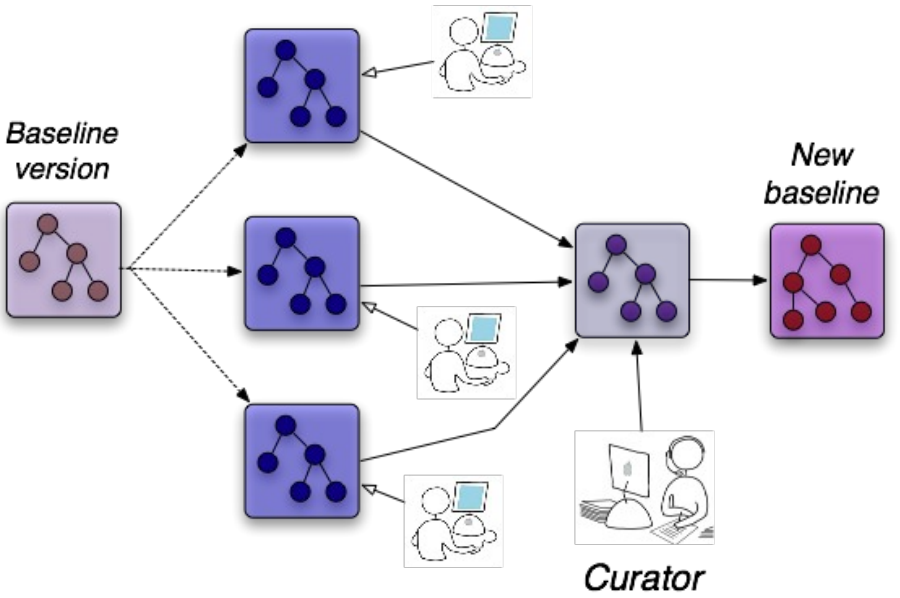


Case II: Parallel Editing, Curation



National Cancer Institute

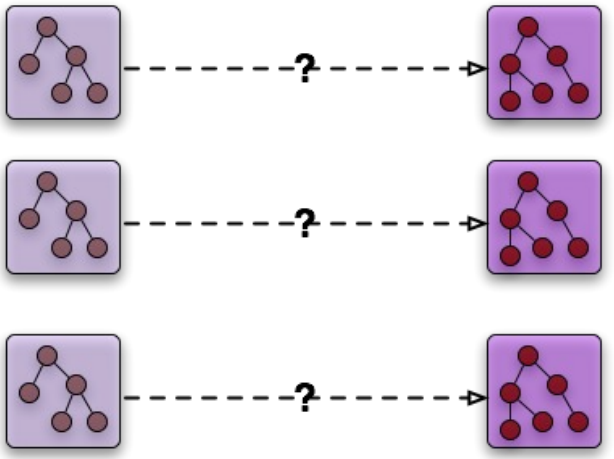
- *Changes are not immediately visible*
- *Need to merge versions and resolve conflicts*
- *Need to accept and reject changes*



Case III: Version Repository

Open Biomedical Ontologies

- *Ontologies developed by different tools in different languages*
- *Usually no record of changes*
- *No record of version compatibility*



Case IV: Enforced workflow

- *Well-defined workflow*
- *Enforced by application*
- *Moderators*
- *Discussion*
- *Voting*

International Classification of Diseases (ICD-10)

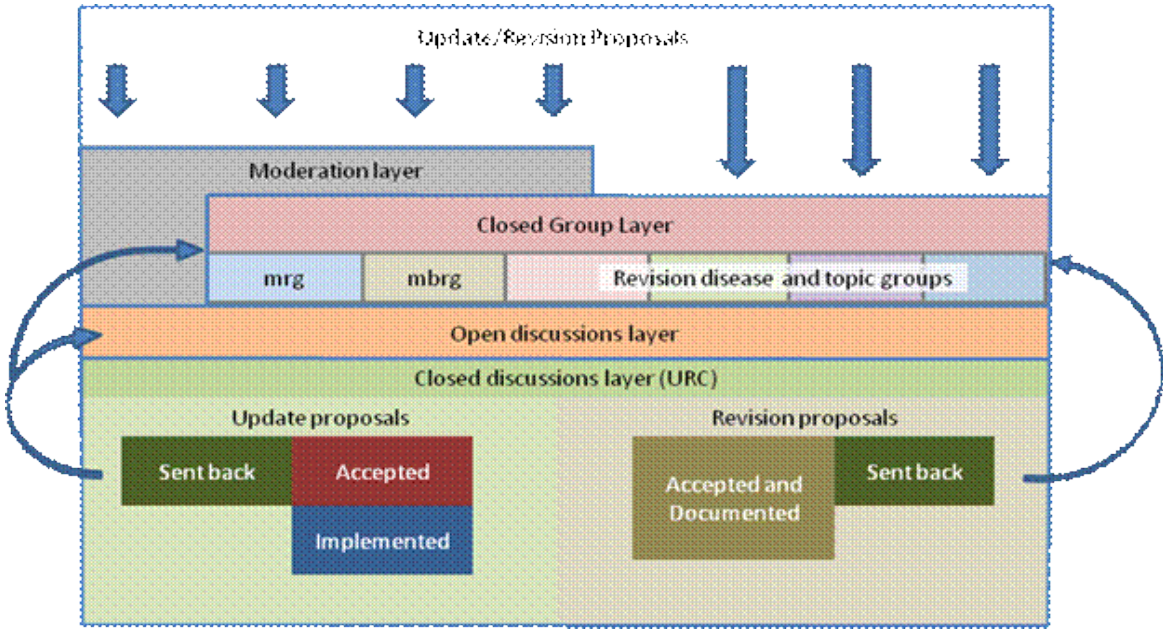
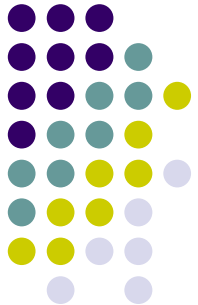


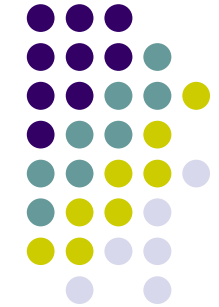
Image Source: http://extranet.who.int/icdrevision/help/docs/UsersGuide_files/image004.gif

Outline



- Introduction and Background
- Tools for collaborative knowledge development
- Use cases
- **Workflows**
- Collaborative Protégé
- Future directions

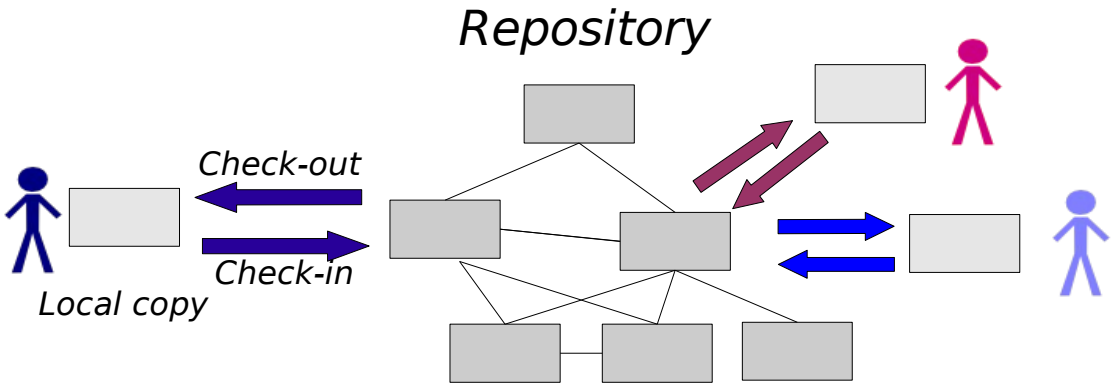
Workflow aspects of collaborative development



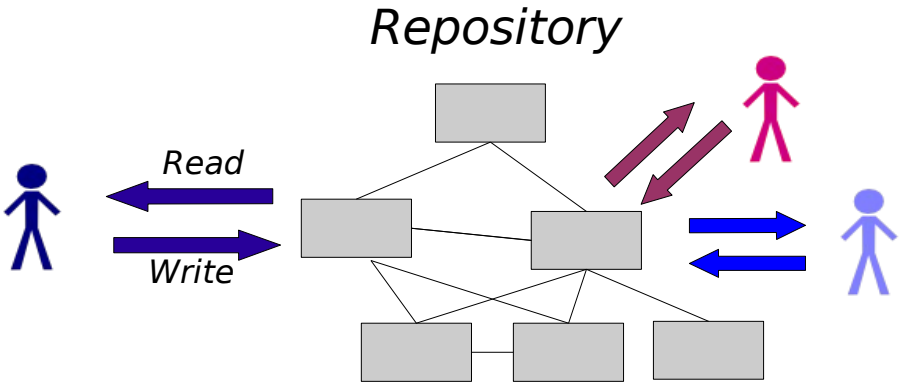
- Versioning
- Simultaneous vs. concurrent development
- Controlled vs. not controlled content
- Access rights and role groups

Versioning

- **File versioning model** (CVS, SVN)
  - check-out and check-in parts of the ontology from a repository
  - lock the checked out part
  - effort at merging changes back in

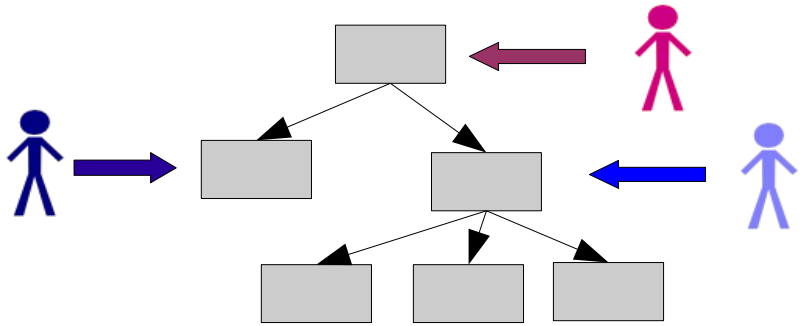


- **Simultaneous access**
  - users edit the same ontology
  - effort at ensuring atomicity of operations

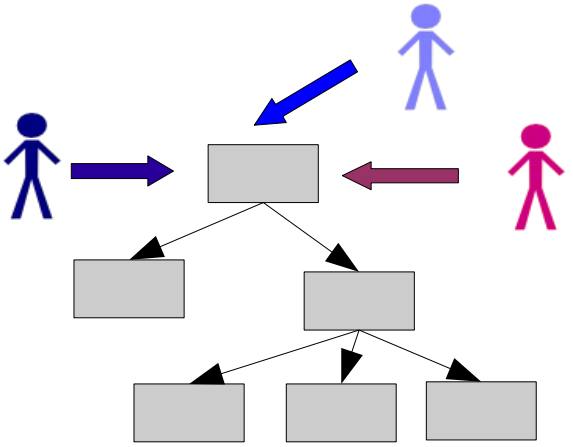


Simultaneous vs. concurrent development

- **Concurrent model**:
  - Split development task in subtasks
  - Each subtask solved by one group
  - Integrate solutions at the end
  - Effort at the end

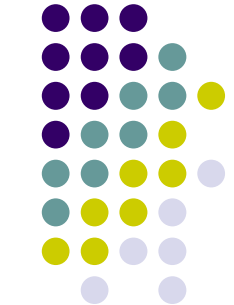


- **Simultaneous model:**
  - Everybody solves the same task (maybe at the same time)
  - Effort throughout the development



- **Hybrid**

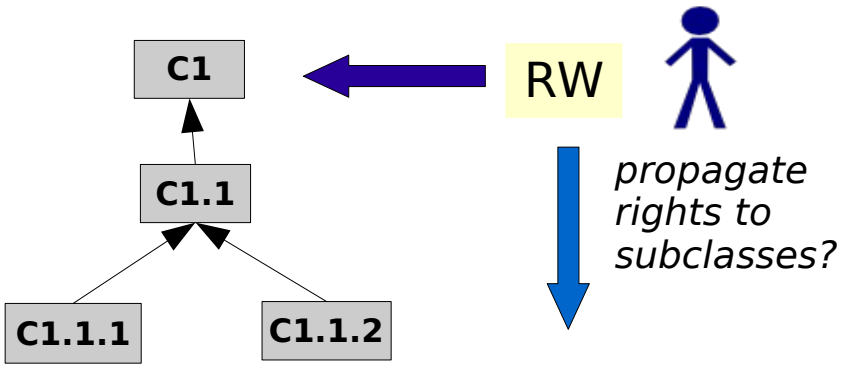
Controlled vs. not controlled content



- No content control
  - Anybody can edit anything at any time
  - Similar to shared tagging in different Internet applications
  - Hard to converge
- Wiki style
  - Anybody can edit anything at any time
  - Wiki “gardeners” to do content clean up
- Accept/reject changes (NCI)
  - Anybody can edit anything at any time
  - Authority that accepts or rejects changes
- Enforced workflow (ICD10)
  - Well defined workflow enforced by the application, e.g. proposals, voting
  - Easier to converge

Access rights and role groups

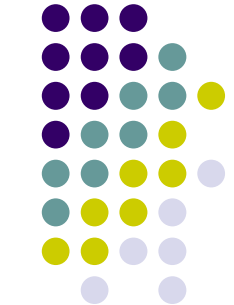
- Different access rights:
  - read
  - write (create, delete, modify)
- Different granularities:
  - on the ontology level
  - on the subtree level
  - on the ontology element level (class, property, individual)
- Need algorithms for computing the access rights for a certain component at a certain time
- Access rights depend on the language semantics -> maybe need inference to compute the actual access rights on an ontology component



RW RW Ø

Pizza **hasTopping** some **PizzaTopping**

Outline

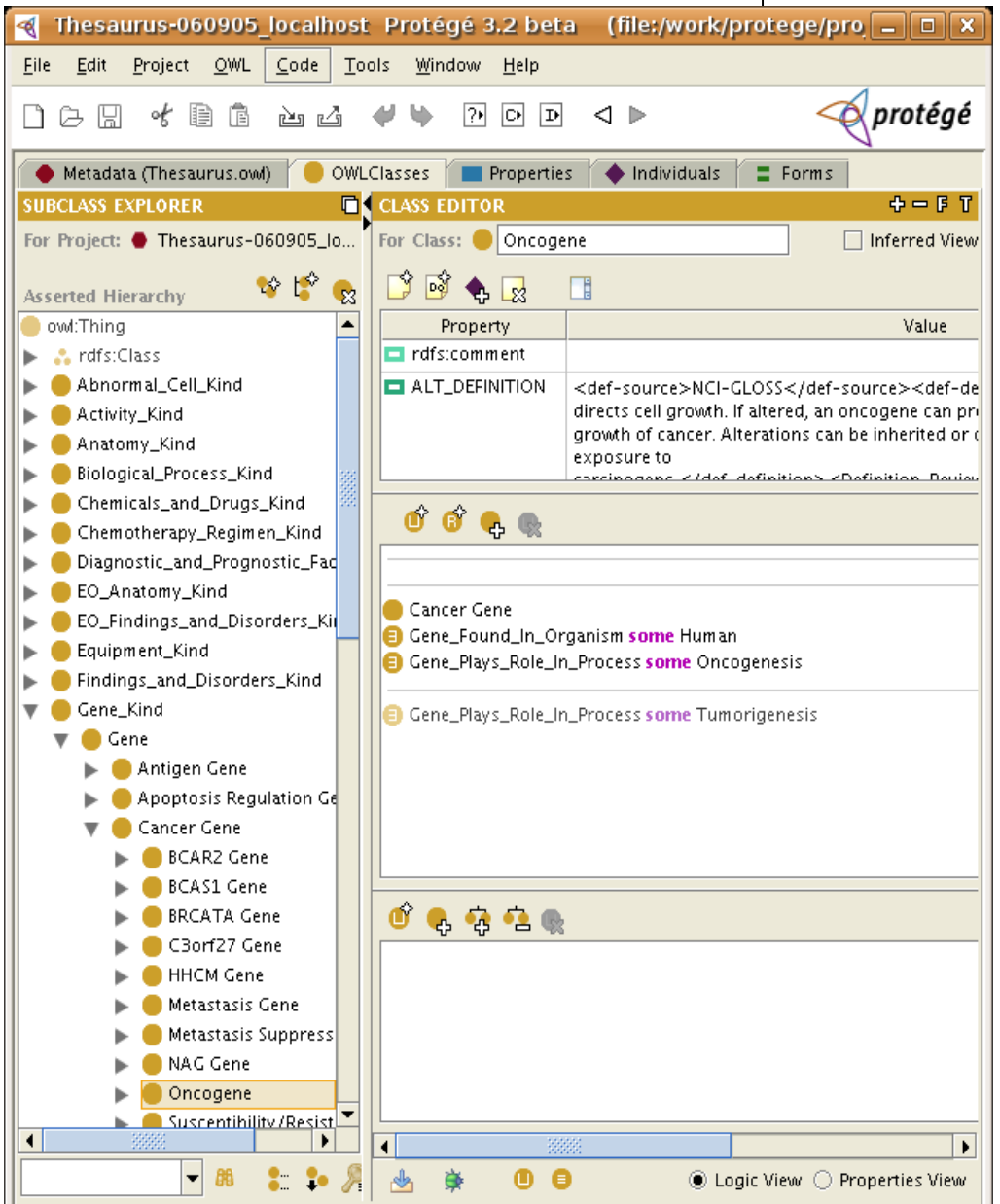


- Introduction and Background
- Tools for collaborative knowledge development
- Use cases
- Workflows
- **Collaborative Protégé**
- Future directions

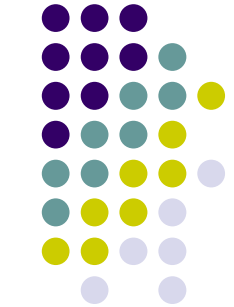
The Protégé ontology editor

- Free, open source ontology editor and knowledge-base framework
- Support for different:
  - ontology languages (OWL, RDF(S), Frames)
  - backends: database, XML, CLIPS, etc.
- Java-based, plug-in architecture
- Strong community: over 80.000 users

http://protege.stanford.edu



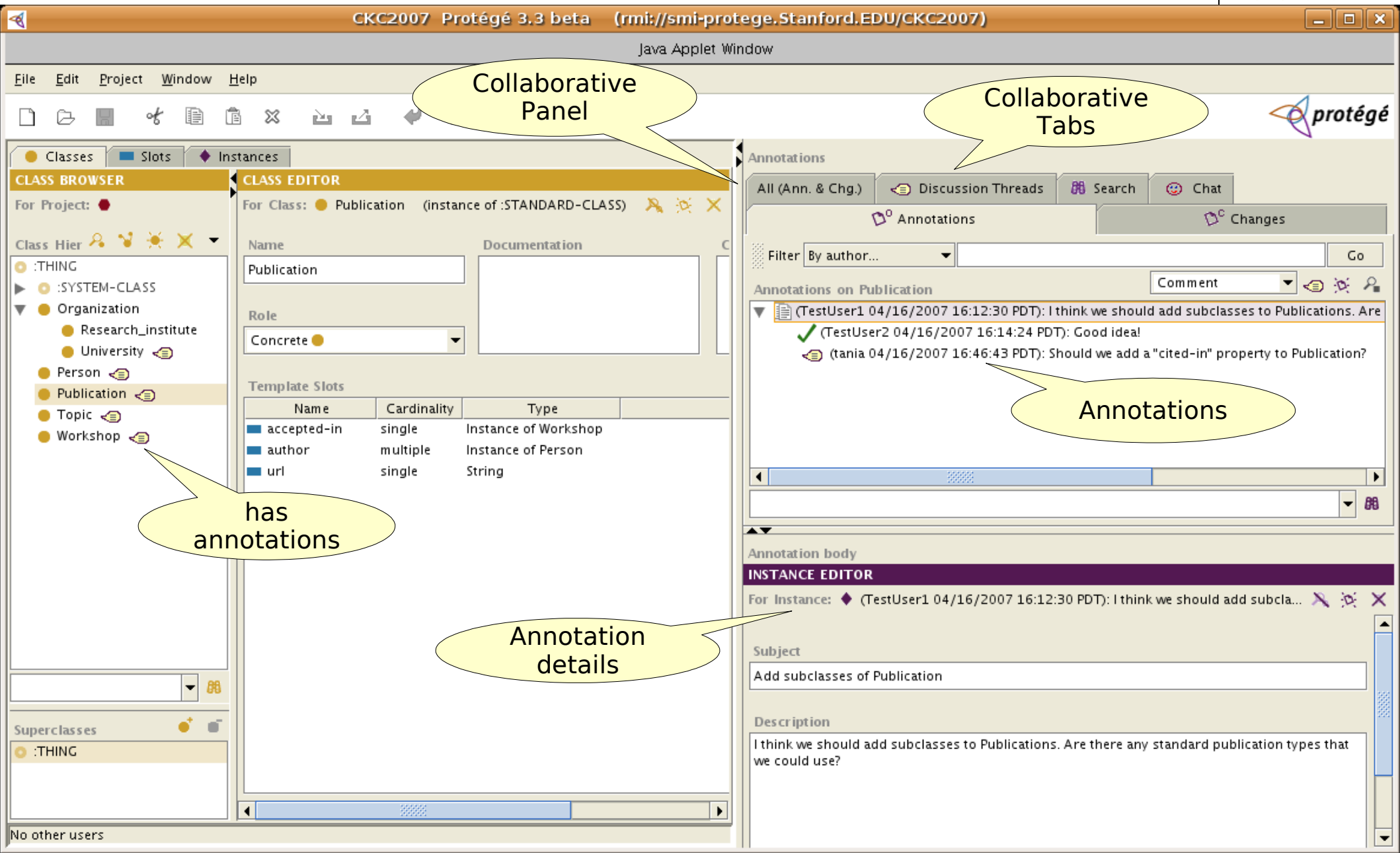
Collaborative Protégé Functionalities



- Extension of existing Protégé system
- Support for:
  - annotating ontology components and changes in the ontology
  - discussion threads
  - proposals and voting
  - searching and filtering
  - defining users, groups, policies
- Works in Protégé OWL and Frames
- Available in multi-user and stand-alone modes
- Distributed with Protégé installation

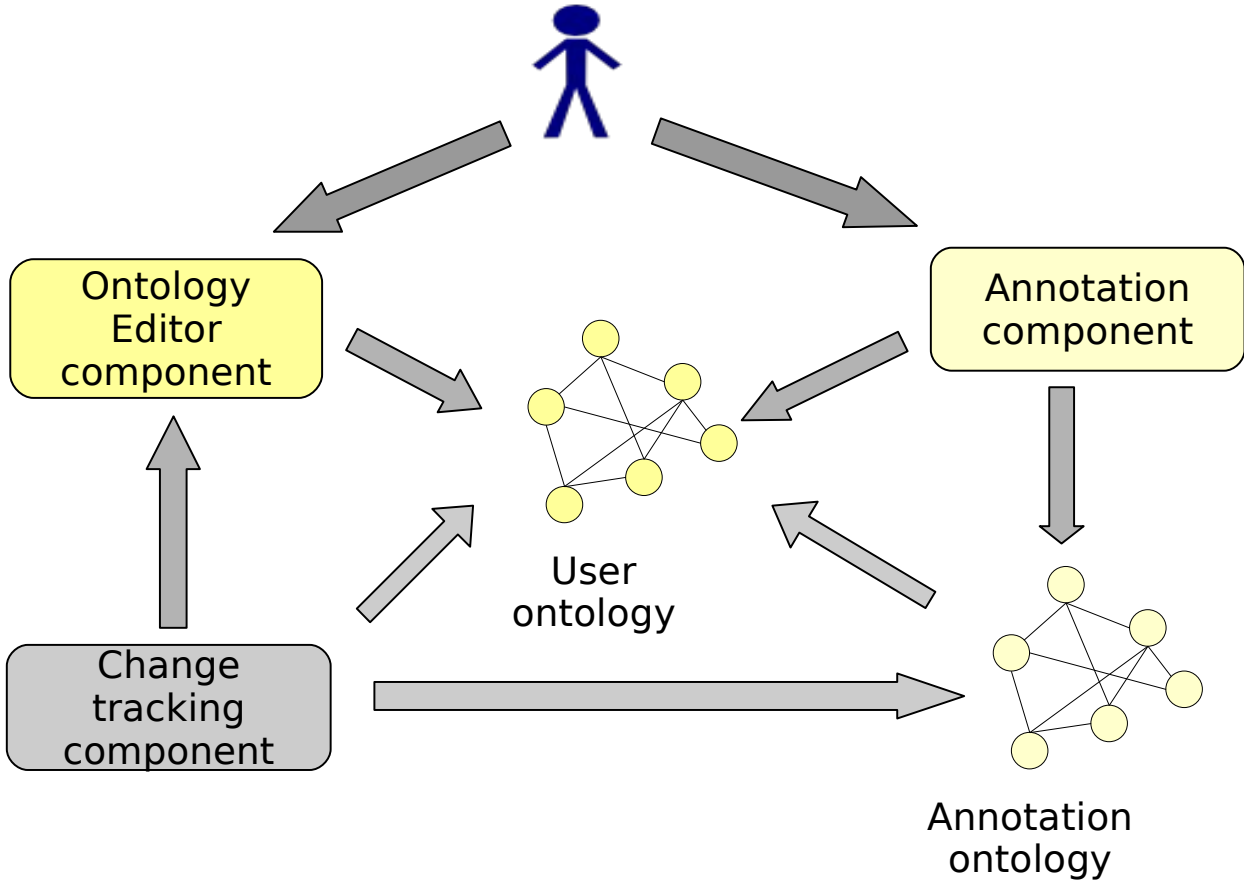
http://protege.stanford.edu/doc/collab-protege/

Collaborative Protégé GUI

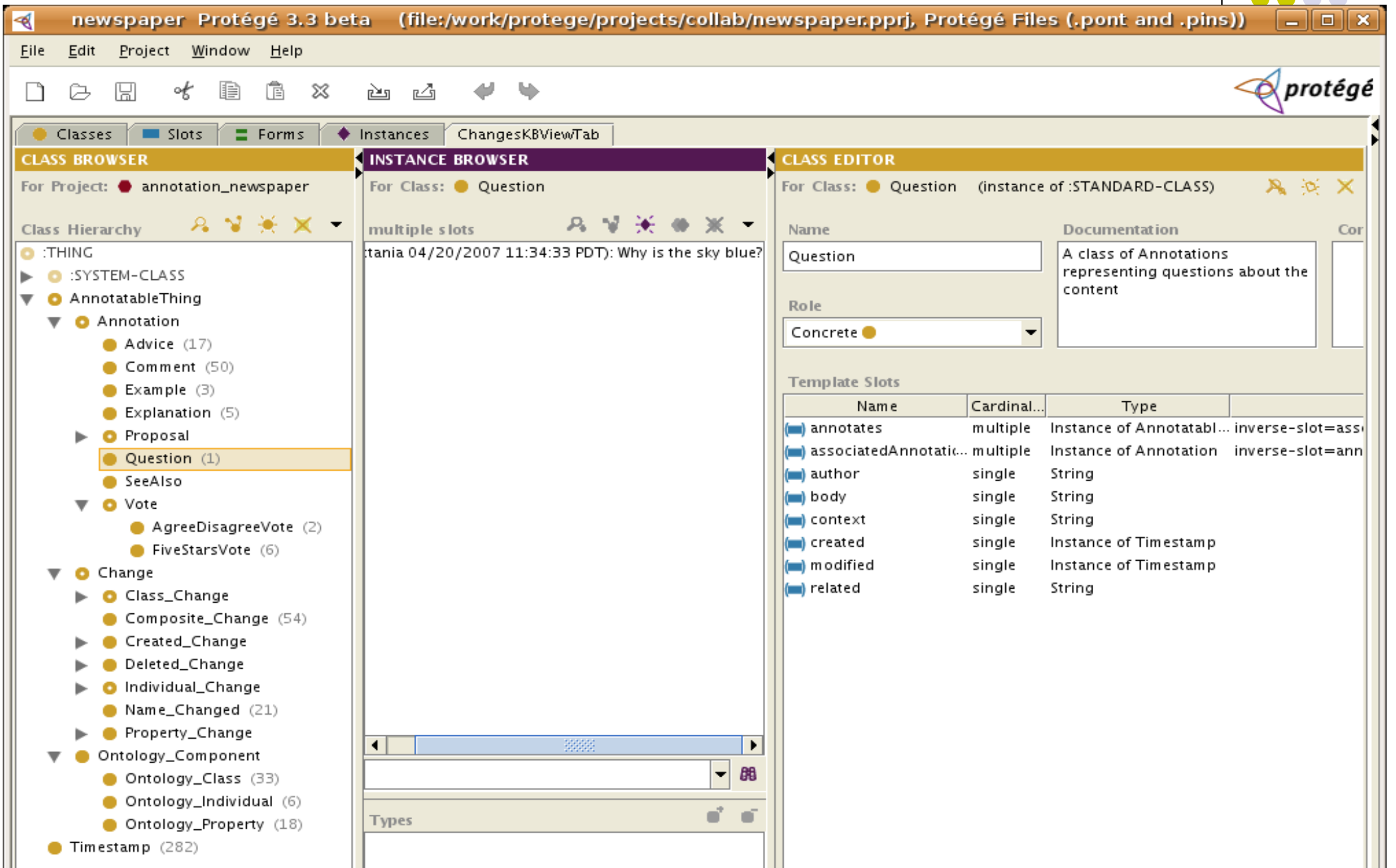


Interactions in Collaborative Protégé

- Ontology editor component:
  - basic ontology editing functionalities
- Annotation component:
  - user ontology is annotated with annotation instances from the Annotation ontology
- Change tracking component:
  - changes are stored as instance of the Annotation ontology

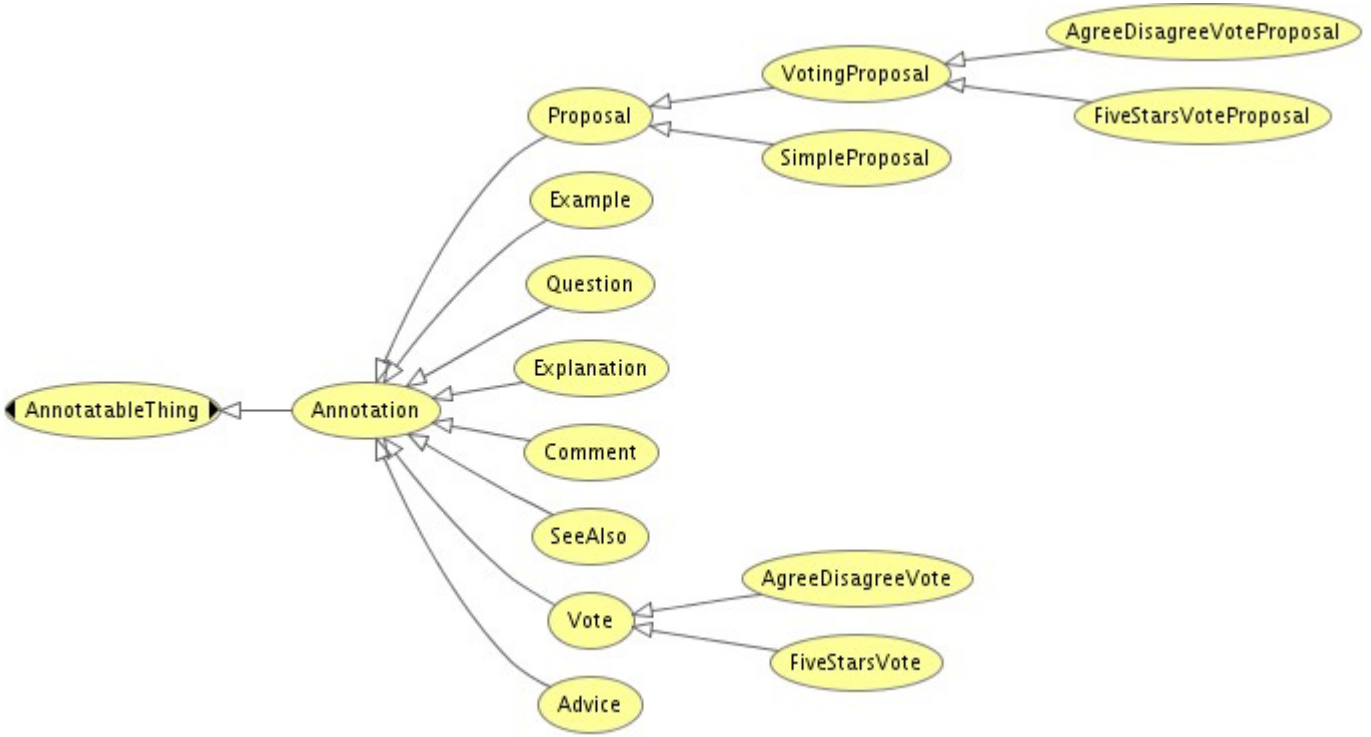


Annotations & Changes Ontology



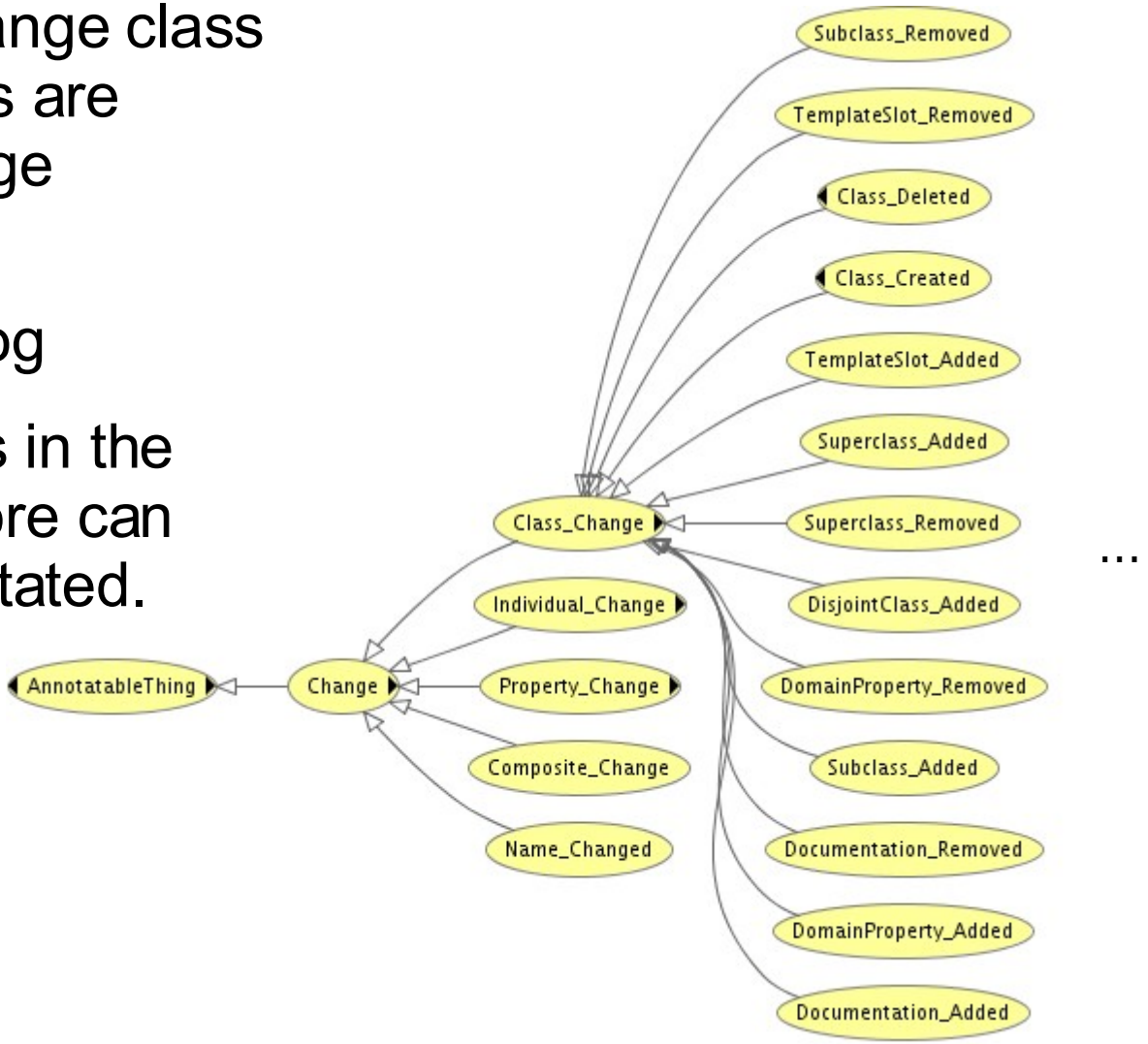
Annotation ontology
Annotation class

- Subclasses of class Annotation provide the annotation types that are available through the Collaborative Protégé UI.



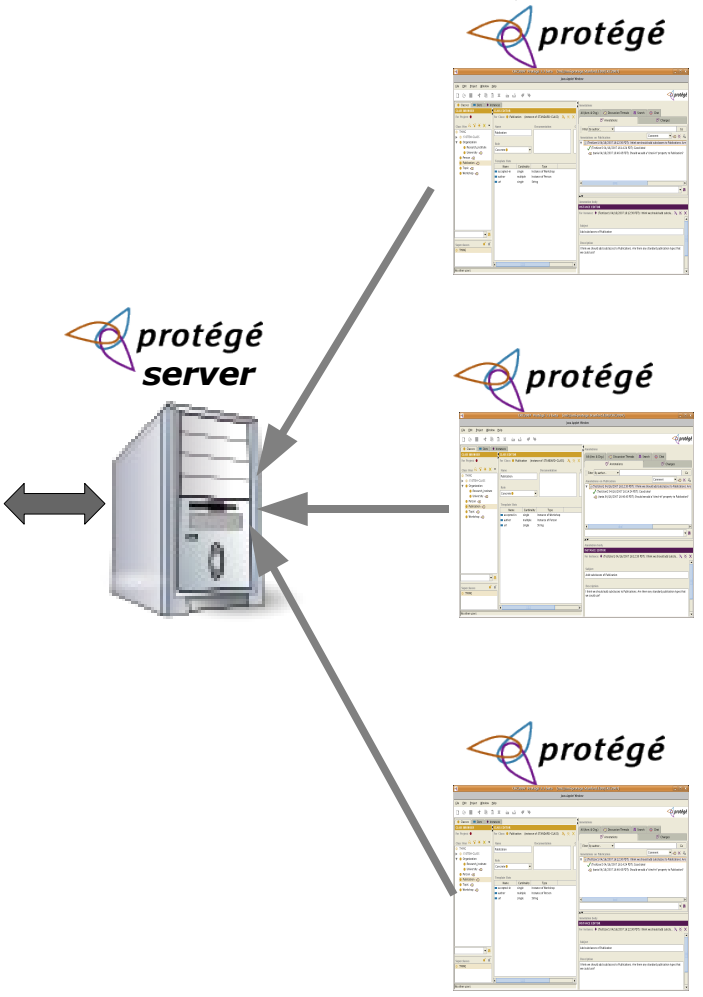
Annotation ontology Change class

- Instances of the Change class and of its subclasses are created by the change tracking component
- Structured change log
- Changes are objects in the ontology and therefore can themselves be annotated.



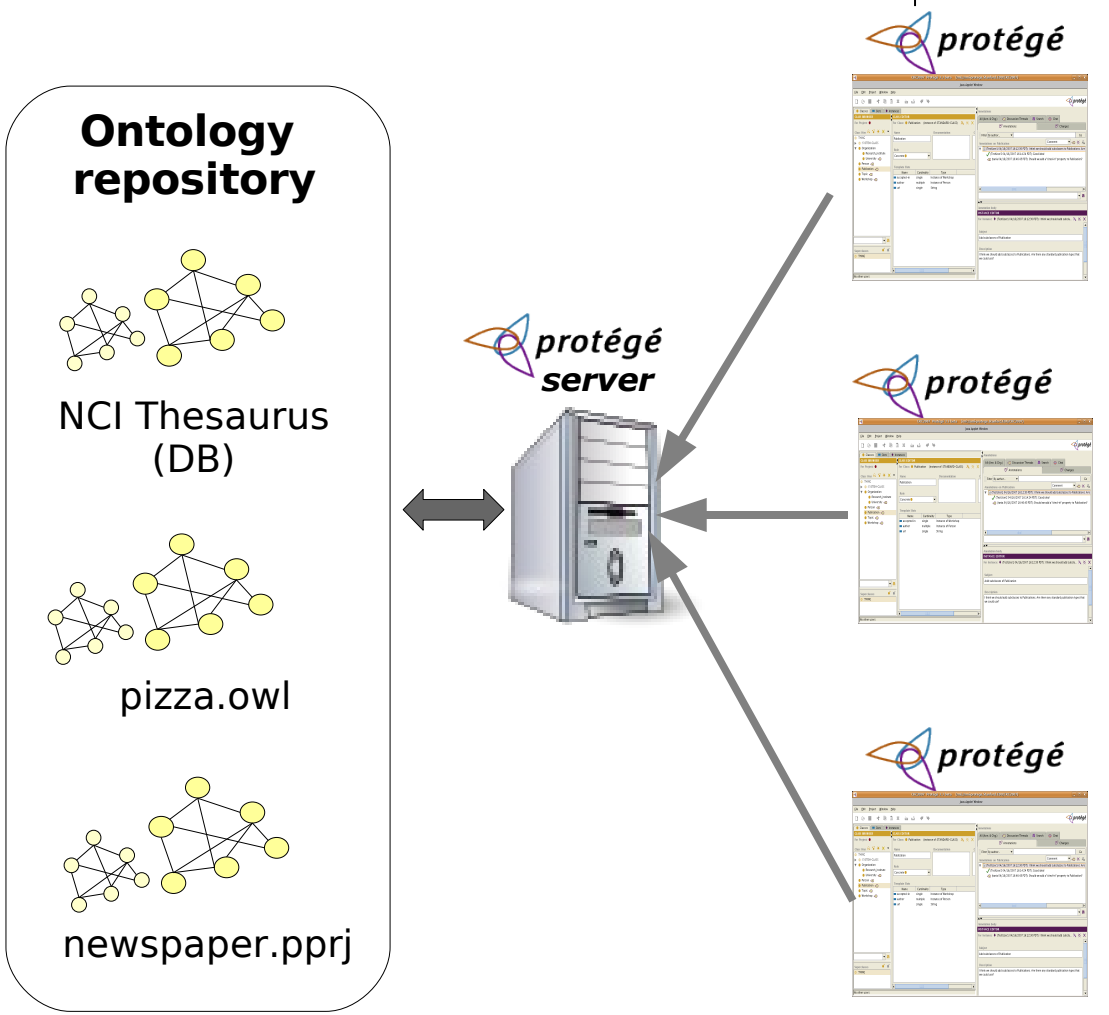
Multi-user Protégé

- Multiple Protégé clients may connect to a Protégé server and edit the same ontology at the same time
- All changes of a client are seen immediately by all other clients
- Configuration of users, groups, policies

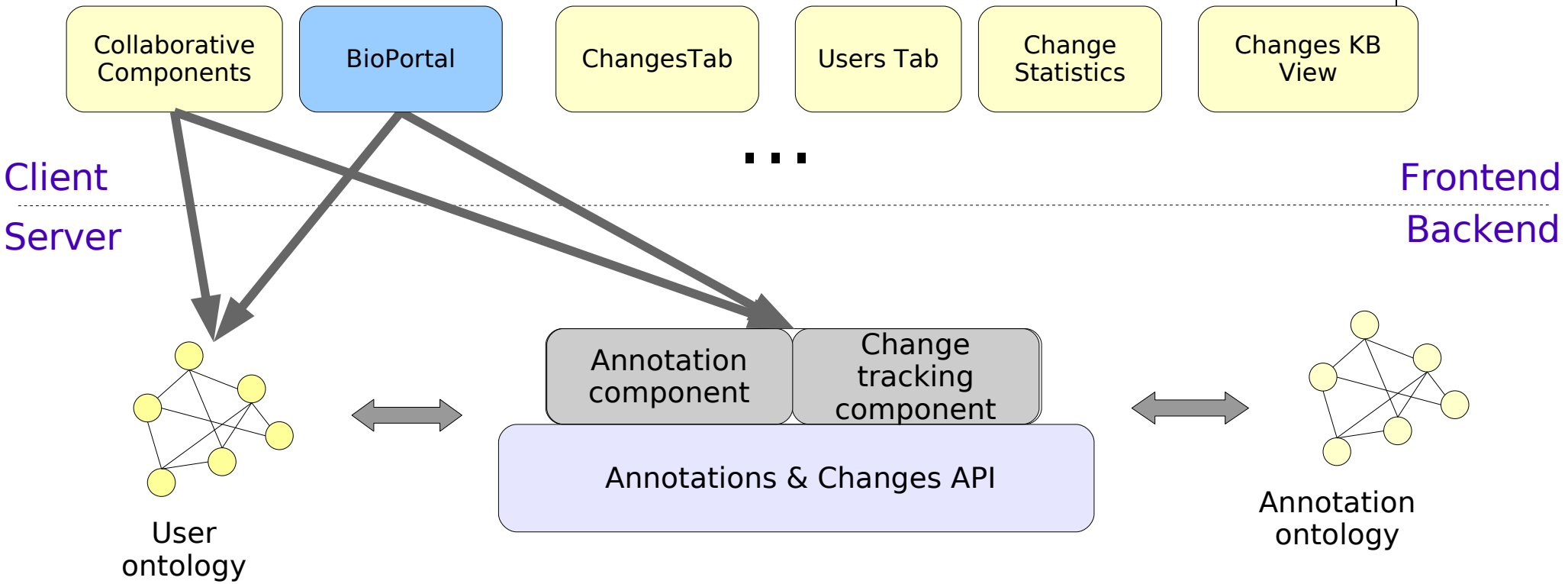


Collaborative Protégé in multi-user mode

- To each ontology on the server we attach an annotation ontology
- All annotations made by a user are seen immediately by other users

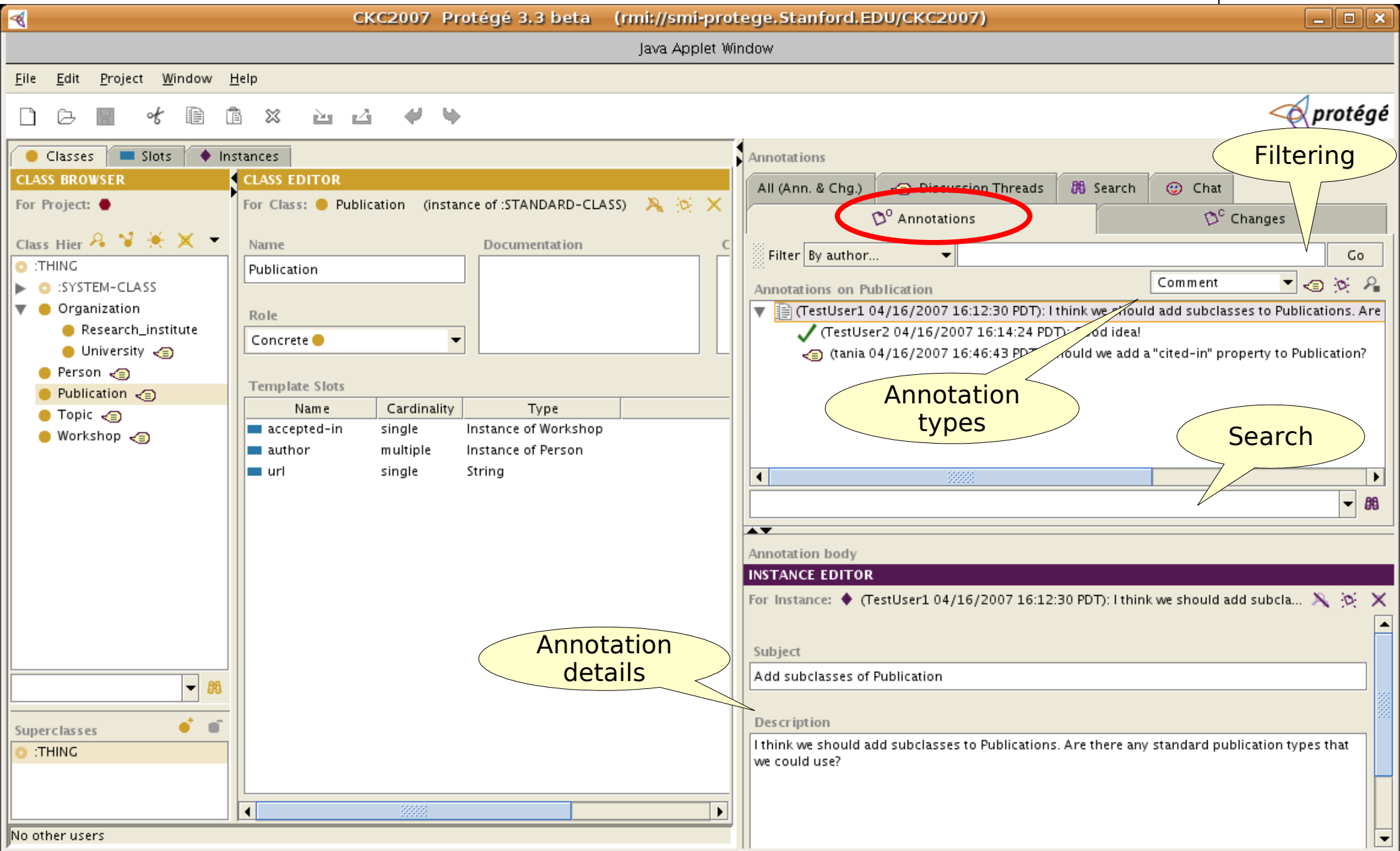


Client-server architecture



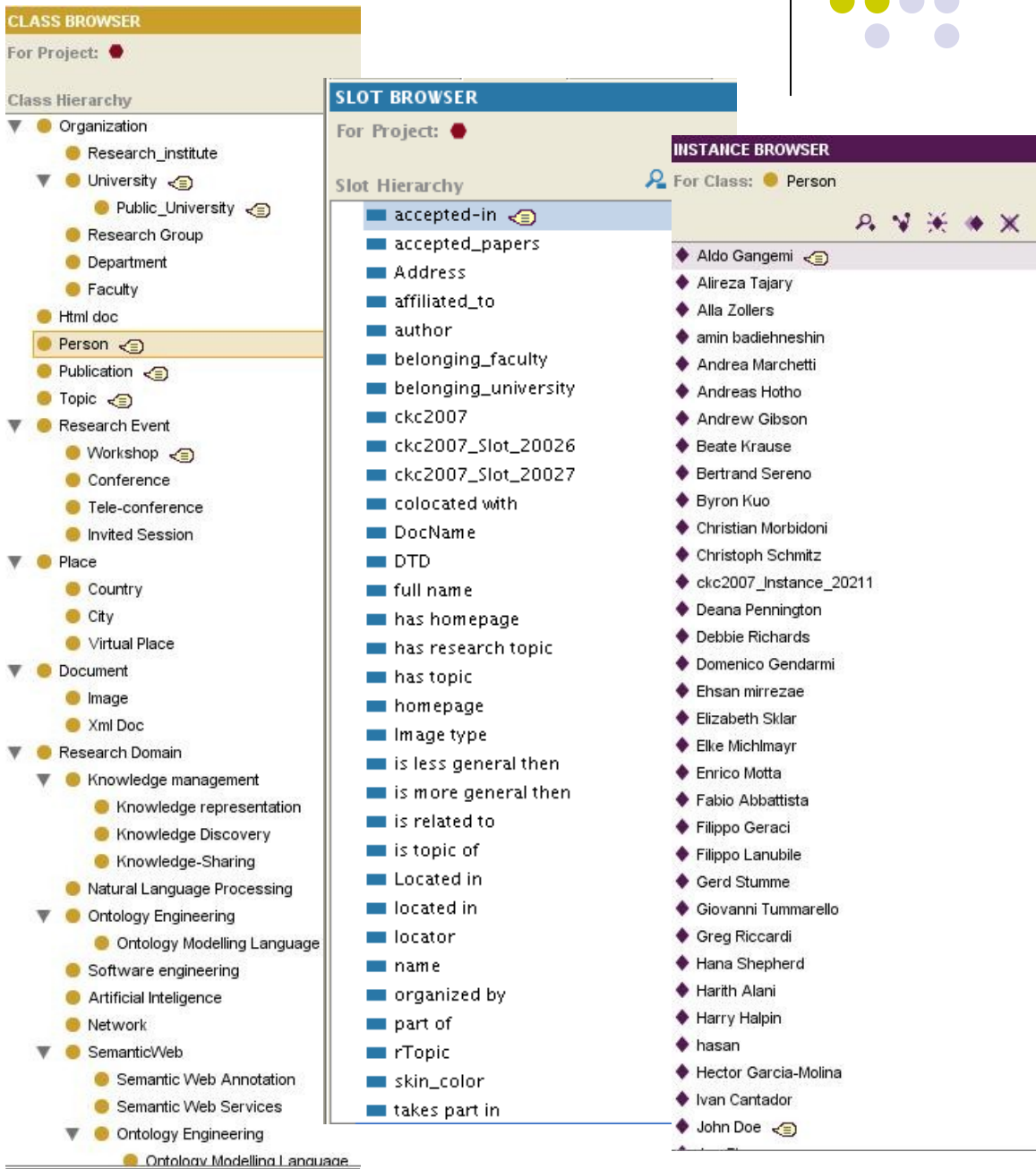
Applications and components on the client side use the common Changes & Annotations API to manipulate the annotations and changes associated to an ontology.

Annotations Tab

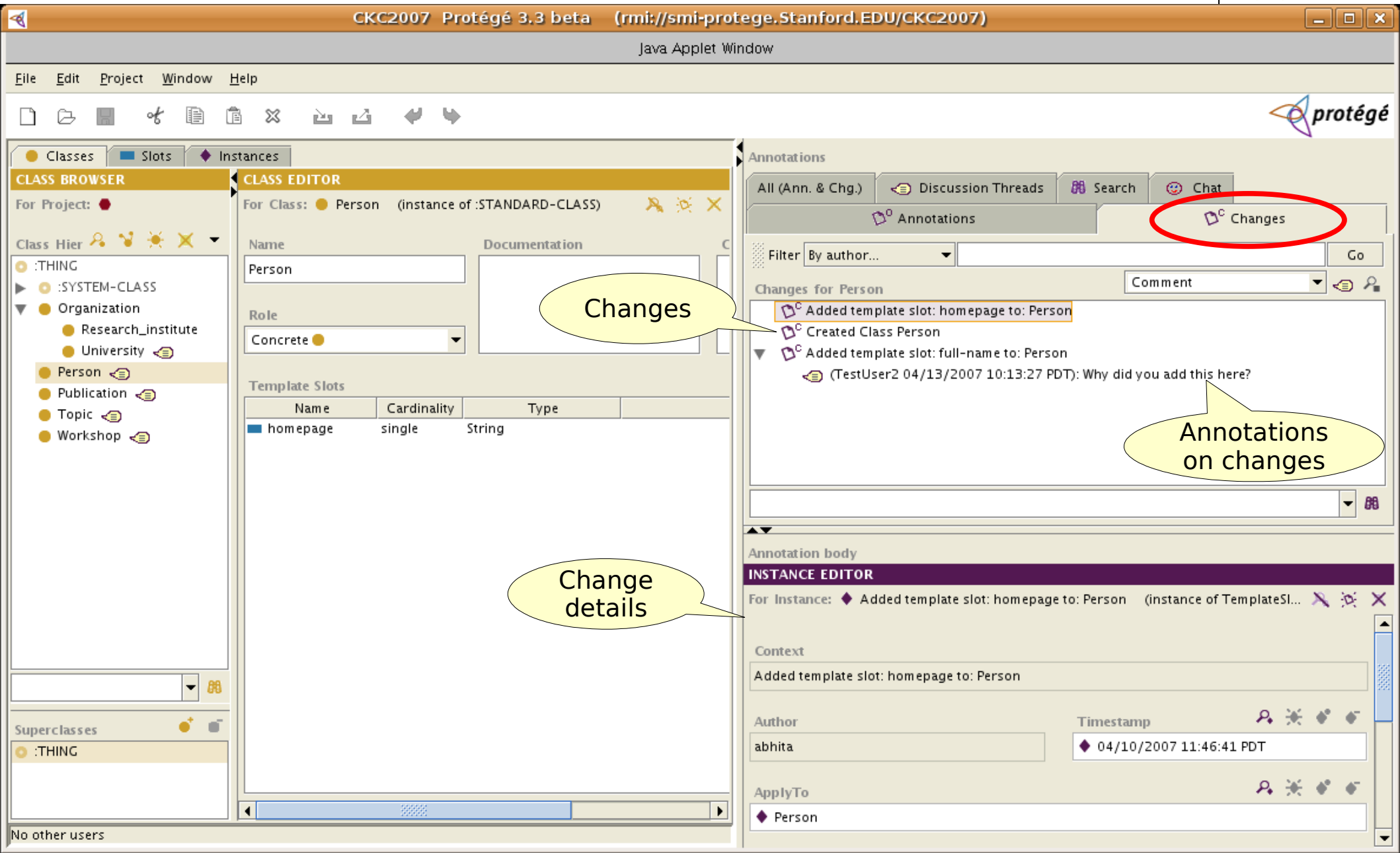


Annotations Tab (cont)

- Annotations are linked to a specific ontology component
- Different types of annotations
- Users may annotate:
  - classes
  - slots (properties)
  - instances (individuals)
- Annotations may be filtered and search based on different criteria



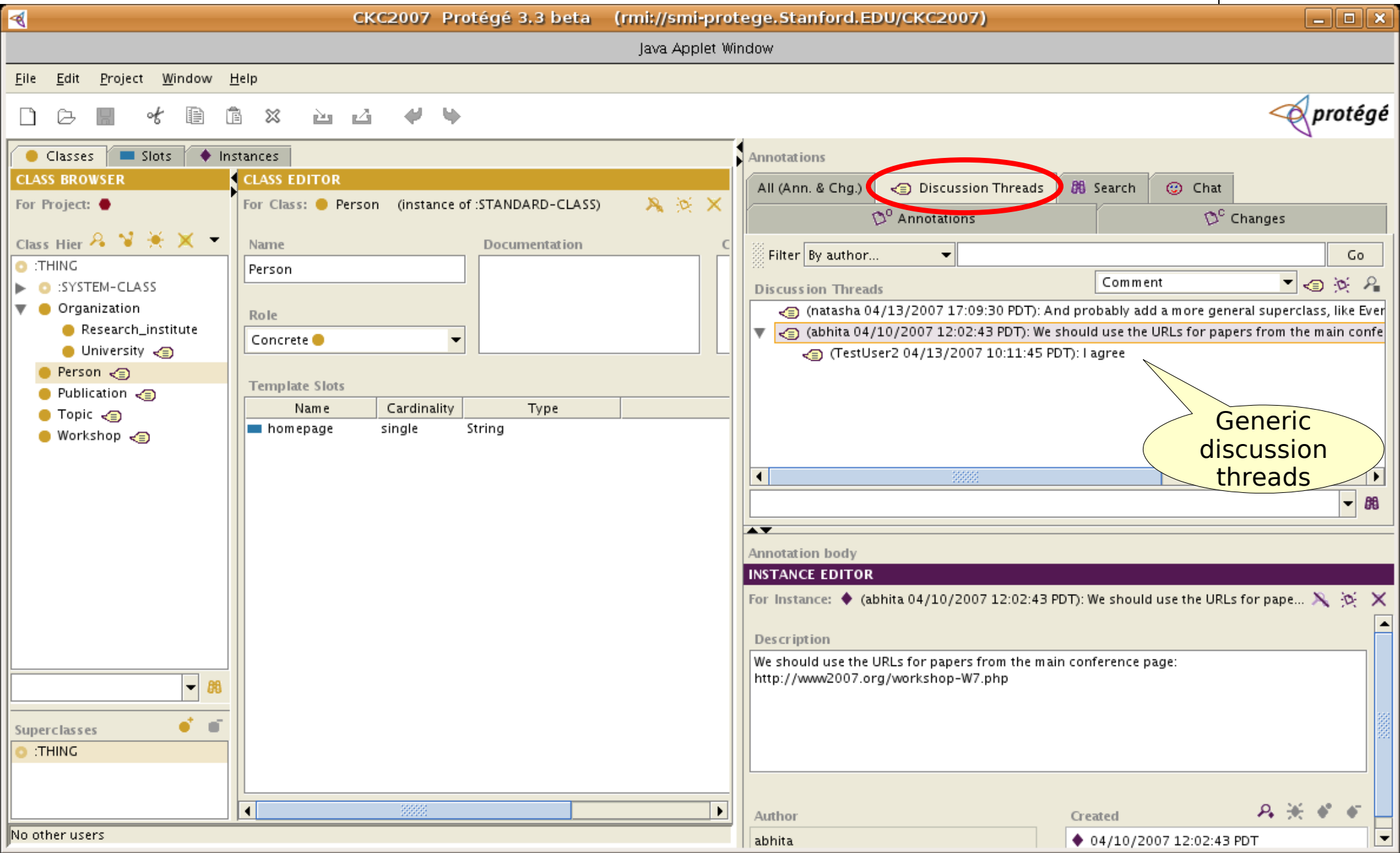
Changes Tab


CKC2007 Protégé 3.3 beta (rmi://smi-protege.Stanford.EDU/CKC2007)
Java Applet Window
File Edit Project Window Help
protégé
Classes Slots Instances
CLASS BROWSER
For Project:
Class Hier
:THING
:SYSTEM-CLASS
Organization
Research_institute
University
Person
Publication
Topic
Workshop
Superclasses
:THING
CLASS EDITOR
For Class: Person (instance of :STANDARD-CLASS)
Name
Person
Documentation
Role
Concrete
Template Slots
Name Cardinality Type
homepage single String
No other users
Annotations
All (Ann. & Chg.)
Discussion Threads
Search
Chat
Annotations
Changes
Filter By author...
Go
Changes for Person
Comment
Added template slot: homepage to: Person
Created Class Person
Added template slot: full-name to: Person
(TestUser2 04/13/2007 10:13:27 PDT): Why did you add this here?
Changes
Annotations on changes
Change details
Annotation body
INSTANCE EDITOR
For Instance: Added template slot: homepage to: Person (instance of TemplateSl...
Context
Added template slot: homepage to: Person
Author
abhita
Timestamp
04/10/2007 11:46:41 PDT
ApplyTo
Person

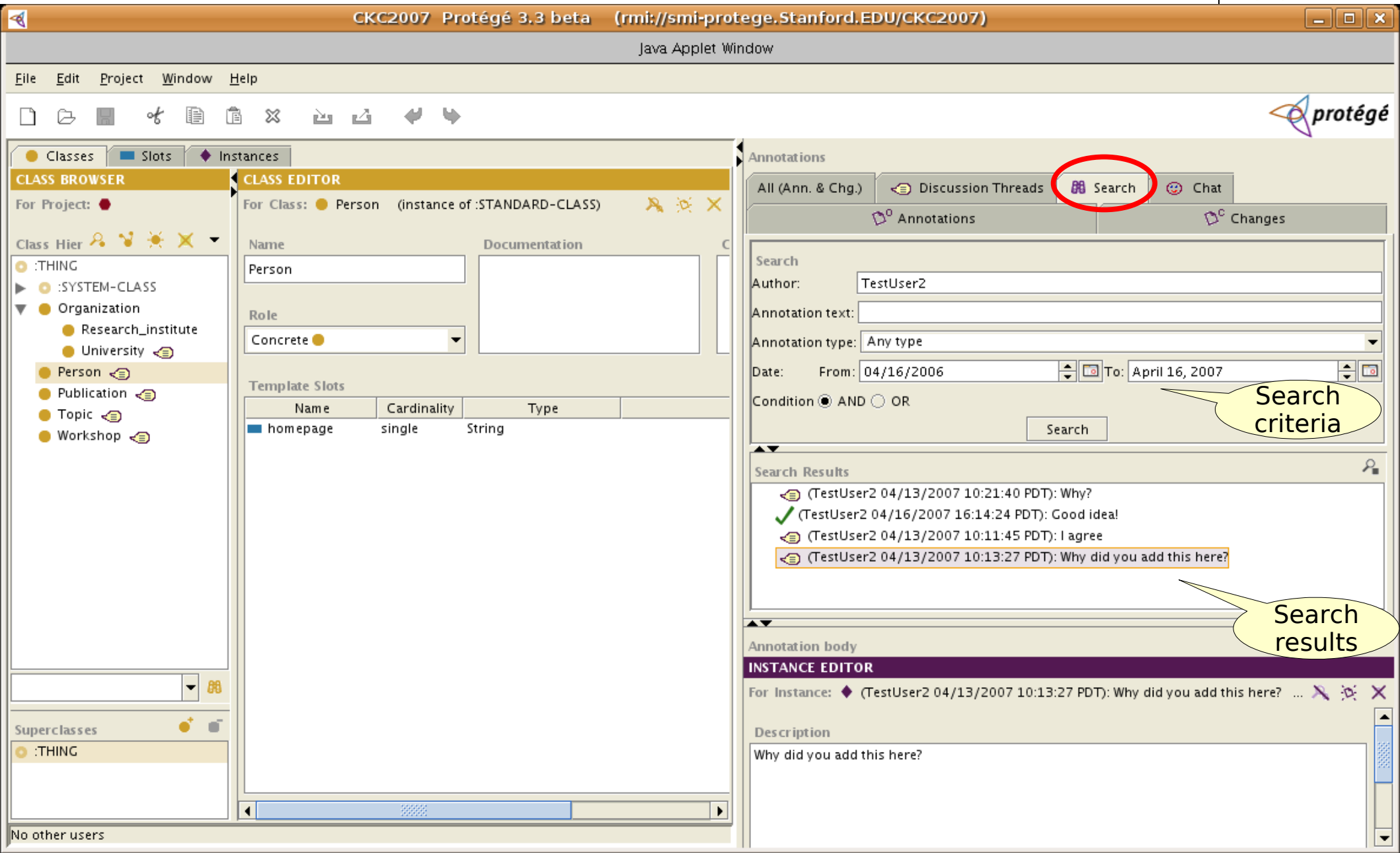
Changes Tab (cont)

- See the history of a concept
- Users may comment on changes; for example on a class rename operation or on a change of a domain property
- Browse the change details (e.g. author, creation date, sub-changes, etc.)

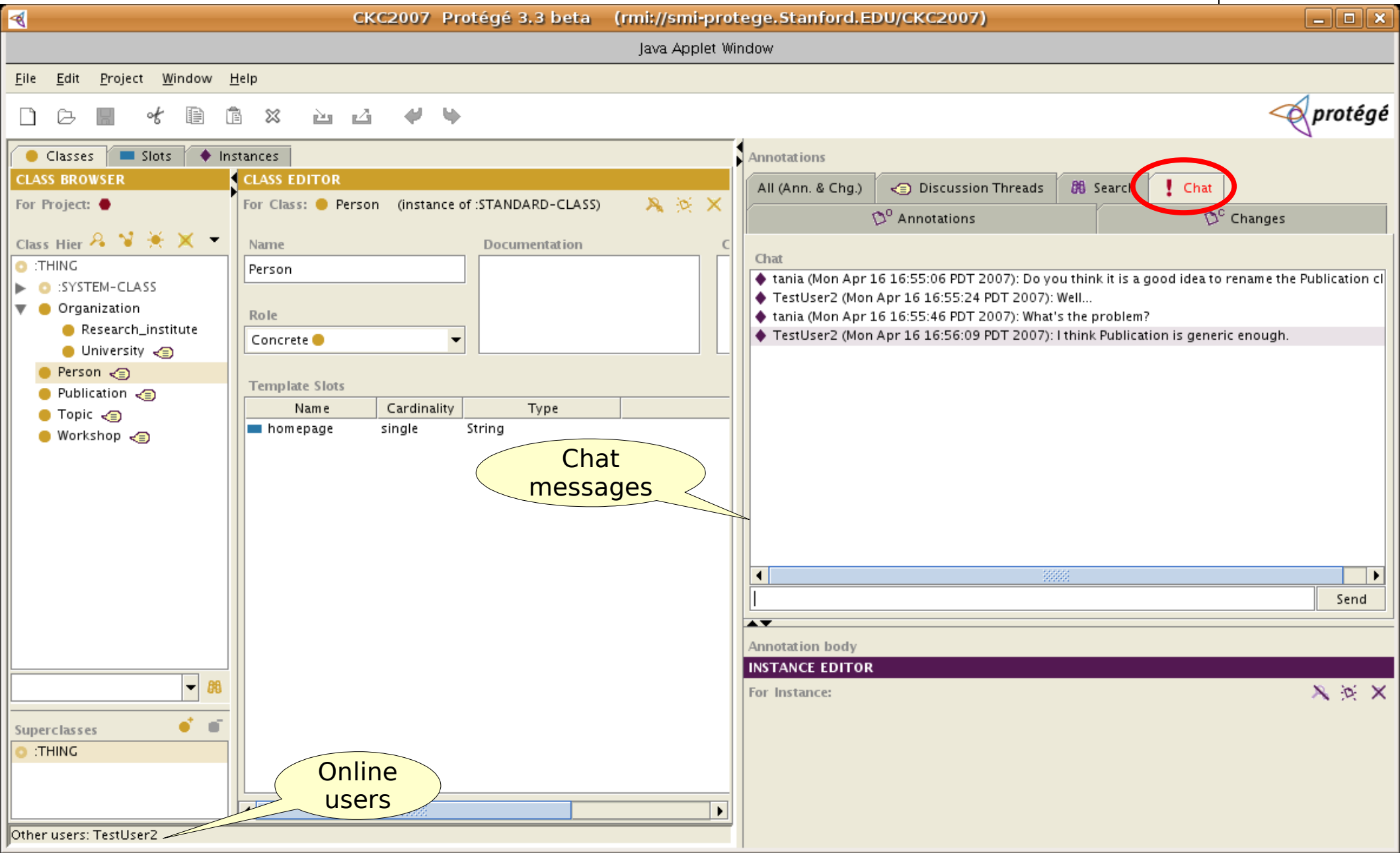
Discussion threads Tab



Search Tab

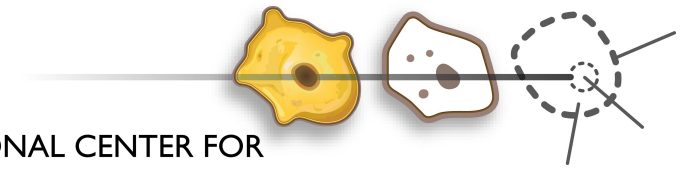

CKC2007 Protégé 3.3 beta (rmi://smi-protege.Stanford.EDU/CKC2007)
Java Applet Window
File Edit Project Window Help
protégé
Classes Slots Instances
CLASS BROWSER
For Project:
Class Hier
:THING
:SYSTEM-CLASS
Organization
Research_institute
University
Person
Publication
Topic
Workshop
Superclasses
:THING
CLASS EDITOR
For Class: Person (instance of :STANDARD-CLASS)
Name
Person
Documentation
Role
Concrete
Template Slots
Name Cardinality Type
homepage single String
No other users
Annotations
All (Ann. & Chg.)
Discussion Threads
Search
Chat
Annotations
Changes
Search
Author: TestUser2
Annotation text:
Annotation type: Any type
Date: From: 04/16/2006 To: April 16, 2007
Condition AND OR
Search
Search criteria
Search Results
(TestUser2 04/13/2007 10:21:40 PDT): Why?
(TestUser2 04/16/2007 16:14:24 PDT): Good idea!
(TestUser2 04/13/2007 10:11:45 PDT): I agree
(TestUser2 04/13/2007 10:13:27 PDT): Why did you add this here?
Search results
Annotation body
INSTANCE EDITOR
For Instance: (TestUser2 04/13/2007 10:13:27 PDT): Why did you add this here?
Description
Why did you add this here?

Chat Tab


CKC2007 Protégé 3.3 beta (rmi://smi-protege.Stanford.EDU/CKC2007)
Java Applet Window
File Edit Project Window Help
protégé
Classes Slots Instances
CLASS BROWSER
For Project:
Class Hier
:THING
:SYSTEM-CLASS
Organization
Research_institute
University
Person
Publication
Topic
Workshop
Superclasses
:THING
CLASS EDITOR
For Class: Person (instance of :STANDARD-CLASS)
Name
Person
Documentation
Role
Concrete
Template Slots
Name Cardinality Type
homepage single String
Annotations
All (Ann. & Chg.)
Discussion Threads
Search
Chat
Annotations
Changes
Chat
tania (Mon Apr 16 16:55:06 PDT 2007): Do you think it is a good idea to rename the Publication cl
TestUser2 (Mon Apr 16 16:55:24 PDT 2007): Well...
tania (Mon Apr 16 16:55:46 PDT 2007): What's the problem?
TestUser2 (Mon Apr 16 16:56:09 PDT 2007): I think Publication is generic enough.
Send
Annotation body
INSTANCE EDITOR
For Instance:
Chat messages
Online users
Other users: TestUser2

The National Center for Biomedical Ontology

- One of three National Centers for Biomedical Computing launched by NIH in 2005
- Collaboration of Stanford, Berkeley, Mayo, Buffalo, Victoria, UCSF, Oregon, and Cambridge
- Primary goal is to make ontologies accessible and usable
- Research will develop technologies for ontology dissemination, indexing, alignment, and peer review
- BioPortal - users may browse, search, visualize ontologies in a web-based portal



NATIONAL CENTER FOR
BIOMEDICAL ONTOLOGY

http://www.bioontology.org/

Bionotes - Windows Internet Explorer

http://localhost:8888/edu.stanford.protegex.bionotes.Bionotes/Bionotes.html

intracellular organelle

Bionotes

THE NATIONAL CENTER FOR BIOMEDICAL ONTOLOGY

BioPortal

Marginal Notes in BioPortal

Ontology: Gene Ontology

Tree View

Search: *intracell* GO!

- ROOT
  - id: oboInOwl:ObsoleteClass
  - molecular_function
  - cellular_component
    - cell part
    - virion part
    - synapse part
    - extracellular matrix part
    - protein complex
    - synapse
    - virion
    - extracellular region part
    - organelle
      - non-membrane-bound organelle
      - intracellular organelle
      - membrane-bound organelle
      - extracellular organelle
      - vesicle
    - membrane-enclosed lumen
    - organelle part
    - extracellular matrix
    - envelope
    - cell
    - extracellular region
  - biological_process
  - rhamnogalacturonan I biosynthetic process
  - modulation by organism of entry into other orga

Class/Type Details

General

Class: intracellular organelle

Id: GO:0043229

Attributes

OBONamespace: cellular_component

Definition: Organized structure of distinctive morphology and function, occurring within the cell. Includes the nucleus, mitochondria, plastids, vacuoles, vesicles, ribosomes and the cytoskeleton. Excludes the plasma membrane.

inSubset: Prokaryotic GO subset

Marginal Notes

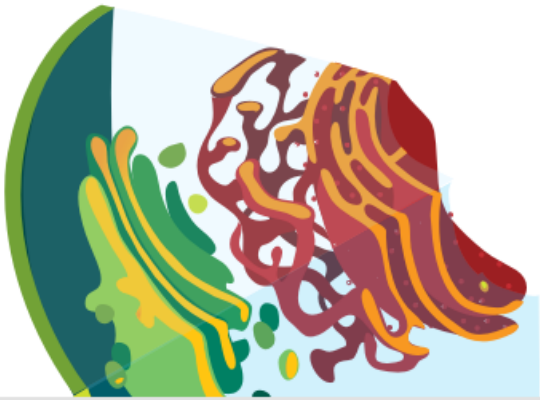
Add Note | Add Reply | Edit Note

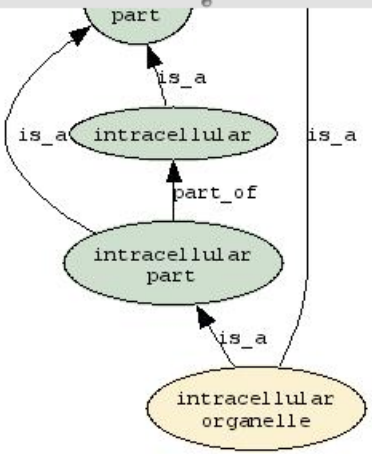
Endoplasmic reticulum

Author:*Dilvan Moreira* Date:*10/03/2007 13:57:05 PDT*

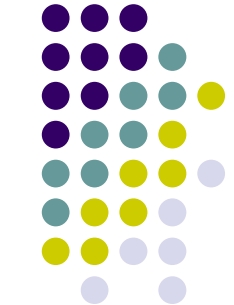
Subject:*Endoplasmic reticulum*

Endoplasmic reticulum and Golgi apparatus in Wikipedia:



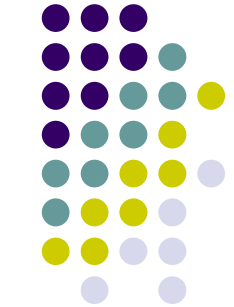


Outline



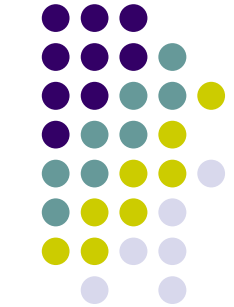
- Introduction and Background
- Tools for collaborative knowledge development
- Use cases
- Workflows
- Collaborative Protégé
- **Future directions**

Future directions
- short term -



- Provide better modularization of the Changes and Annotations API
- Plugin architecture – other developers may add their own collaborative tabs
- Make the UI configurable: for example, see only the annotations and the search tab
- Chat with hyperlinks to ontology concepts
- Easier setting up of the collaborative features
- Optimizations with respect to performance and scalability

Future directions (cont)
- longer term -

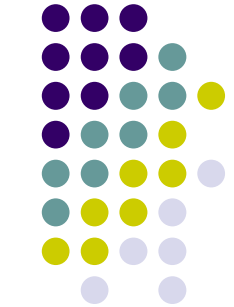


- Support for different workflow models
- Porting the collaborative components to Protégé 4
- Integrate into WebProtege

Feedback welcome!

tudorache@stanford.edu

References



[1] Collaboration: http://collab.blueoxen.net/cgi-bin/wiki.pl?Collaboration

[2] CKC Challenge: http://km.aifb.uni-karlsruhe.de/ws/ckc2007/challenge.html

[3] Tech Report with challenge results and discussions:

http://bmir.stanford.edu/publications/view.php/the_ckc_challenge_exploring_tools_for_collaborative_knowledge

CKC Challenge Tools:

[4] http://www.bibsonomy.org/group/ckc2007

[5] http://protege.stanford.edu/doc/collab-protege/

[6] http://www.dbin.org/

[7] http://www.hozo.jp/

[8] http://soboleo.fzi.de:8080/webPortal/

BioPortal:

[9] http://www.bioontology.org/ncbo/faces/index.xhtml